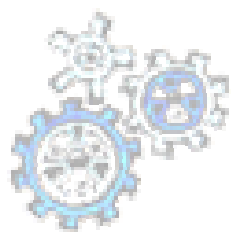
Privacy in Mobile Networks

Maurizio Atzori

atzori@di.unipi.it

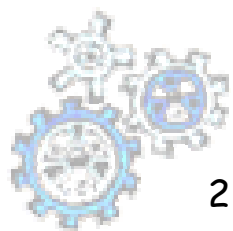
KDDLab, ISTI-CNR and

CS Dept., University of Pisa



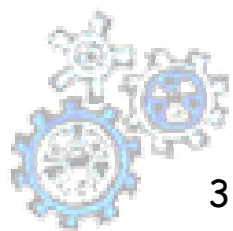
Summary

- Privacy Requirements in Mobile Systems
- Brief survey of Privacy in Networks
  - Anonymous Routing
  - Anonymous Web
- Location Privacy for Location Based Services (LBSs)
- Conclusions



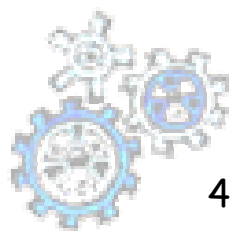
Privacy Requirements

- Content Privacy
  - Mainly security, cryptography
- Identification Privacy
  - Private from unauthorised users
- Location Privacy
  - Mobile user should be untraceable



Anonymous Routing

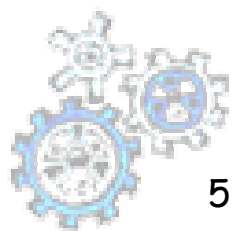
- DC-Nets
  - Based on secure multiparty computation, high complexity $O(n^3)$
- Mix-Nets and Onion Routing
  - By David Chaum (since 1981!)
  - Nick Mathewson implemented Tor and Mixminion
- Crowds
  - Paths change randomly
- CliqueNet
  - Based on small DC-Nets, more scalable but less privacy



Sketch of Mix-Nets and Onion Routing Protocols

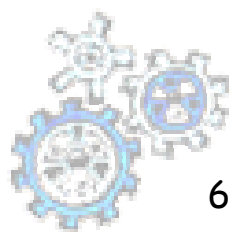
- Alice wants to send the message M to Bob
  - Alice sends $E_{S1}(S2, E_{S2}(Bob, E_{Bob}(M)))$ to S1
  - S1 decrypts the message obtaining
    - S2 and $E_{S2}(Bob, E_{Bob}(M))$
  - S1 sends $E_{S2}(Bob, E_{Bob}(M))$ to S2
  - S2 decrypts the message obtaining
    - Bob and $E_{Bob}(M)$
  - S2 sends $E_{Bob}(M)$ to Bob

- Every communication should be untraceable:
  - Delayed until there are several messages, Messages of fixed length, several fake messages



Anonymous Publishing (FreeNet)

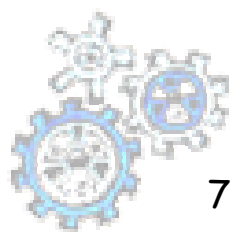
- It is possible to publish web pages anonymously
  - Redundancy to avoid censury
  - High latency
- Also surfing is anonymous
  - It is impossible to trace the path of FreeNet users
- The concept of Darknets
  - Microsoft (2002) ACM Workshop on Digital Rights Management



K-Anonymous Message Transmission

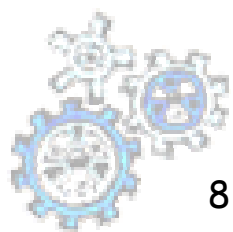
- Total anonymity
  - No information about senders or receivers
- Sender k-anonymity
  - Given a message, the attacker can only narrow down its search to a set of (at least) k senders
- Receiver k-anonymity
  - Same for receivers

- A formal model but no implementation yet
  - Probably high latency... Feasible approach?



Personalized Anonymization for Location Privacy

- Anonymity for protecting location privacy
- Context: communication for Location-based services (LBS)
  - Trusted Anonymization Server between user and LBS
- CliqueCloak Algorithm
  - Mask location and temporal data by perturbation
  - Based on delaying messages and lowering the spatio/temporal resolution
- Each user can specify herown parameters
  - K, QoS (Space Resolution, Time Precision)

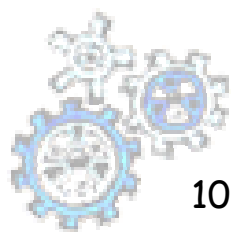


K-Anonymization

- Anonymity: "*a state of being not identifiable within a set of subjects, the Anonymity Set*"
- K-Anonymity: |Anonymity Set| ≥ k
- Subjects of the data cannot be re-identified while the data remain practically useful
  - By attribute generalization and tuple suppression

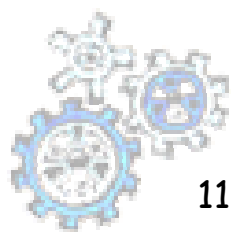
Original Database

Race	DOB	Sex	ZIP	Problem
black	05/20/1965	M	02141	short of breath
black	08/31/1965	M	02141	chest pain
black	10/28/1965	F	02138	painful eye
black	09/30/1965	F	02138	wheezing
black	07/07/1964	F	02138	obesity
black	11/05/1964	F	02138	chest pain
white	11/28/1964	M	02138	short of breath
white	07/22/1965	F	02139	hypertension
white	08/24/1964	M	02139	obesity
white	05/30/1964	M	02139	fever
white	02/16/1967	M	02138	vomiting
white	10/10/1967	M	02138	back pain



2-anonymized Database

Race	DOB	Sex	ZIP	Problem
black	1965	M	02141	short of breath
black	1965	M	02141	chest pain
black	1965	F	02138	painful eye
black	1965	F	02138	wheezing
black	1964	F	02138	obesity
black	1964	F	02138	chest pain
white	196*	*	021**	short of breath
white	196*	*	021**	hypertension
white	1964	M	02139	obesity
white	1964	M	02139	fever
white	1967	M	02138	vomiting
white	1967	M	02138	back pain

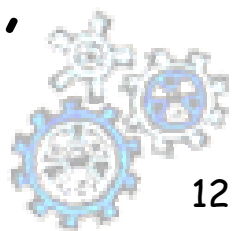


Messages

- ms = <uid, rno, {t,x,y}, k, {dt, dx, dy}, C>

- Where
  - (uid, rno) = user-id and message number
  - {t,x,y} = L(ms) = spatio-temporal location
  - K = anonymity threshold
  - dt, dx, dy = quality of service constraints
  - C = the actual message

  - Bcn(ms) = [t-dt, t+dt] [x-dx, x+dx] [y-dy, y+dy]
  - Bcl(ms) = spatio-temporal cloaking box of ms, contained in Bcl(ms)

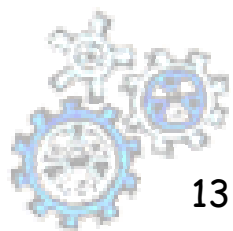


Definition of Location k-anonymity

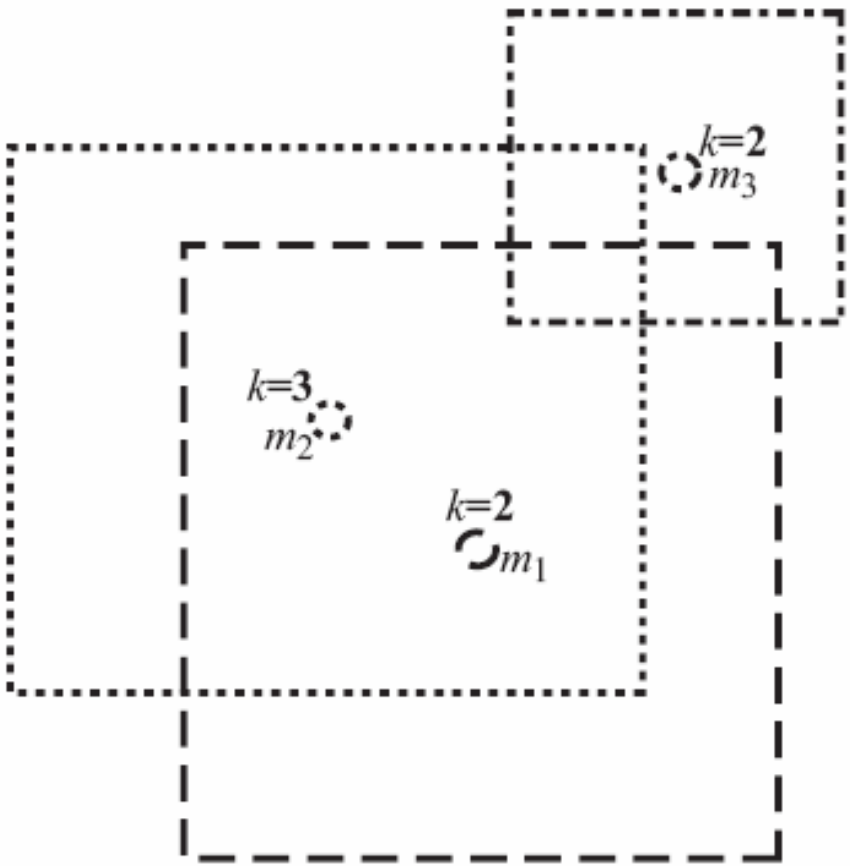
- For a message ms in S and its perturbed format mt in T, the following condition must hold:

$$\forall\ T' \subset T,\ \text{s.t. } mt \in T',\ |T'| \geq ms.k,$$
$$\forall\ \{mt_i, mt_j\} \subset T',\ mt_i.uid \neq mt_j.uid \text{ and}$$
$$\forall\ mti \in T',\ Bcl(mt_i) = Bcl(mt)$$

- $ms.C = mt.C$, $mt.uid = hash(ms.uid)$

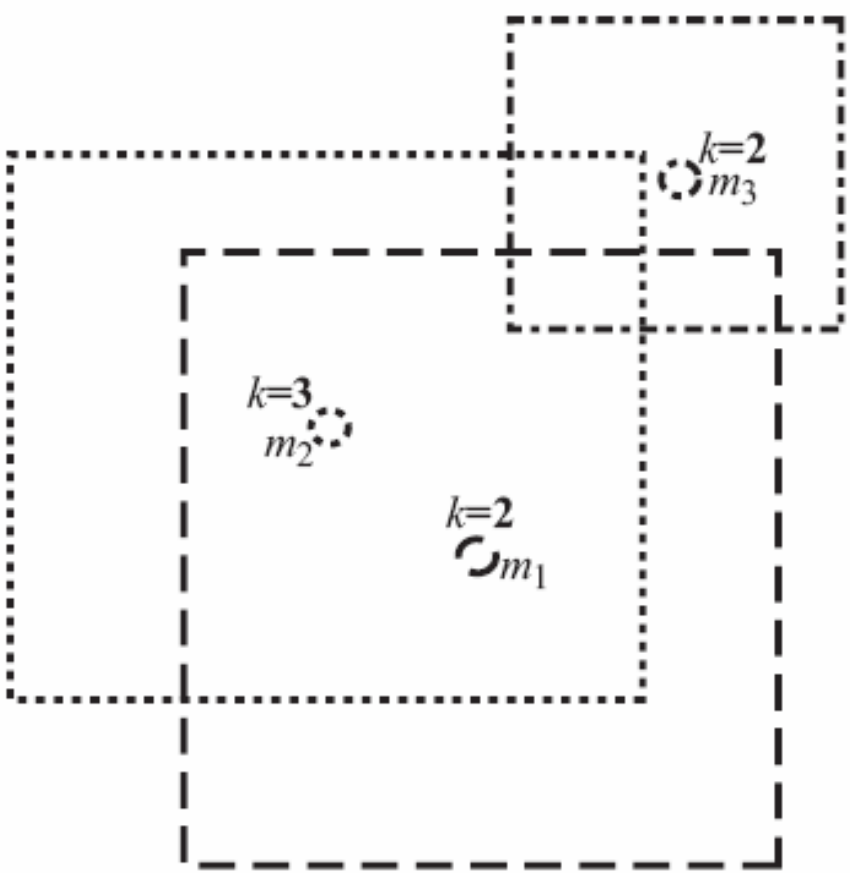


Clique-Cloak Algorithm: Spatial Layouts

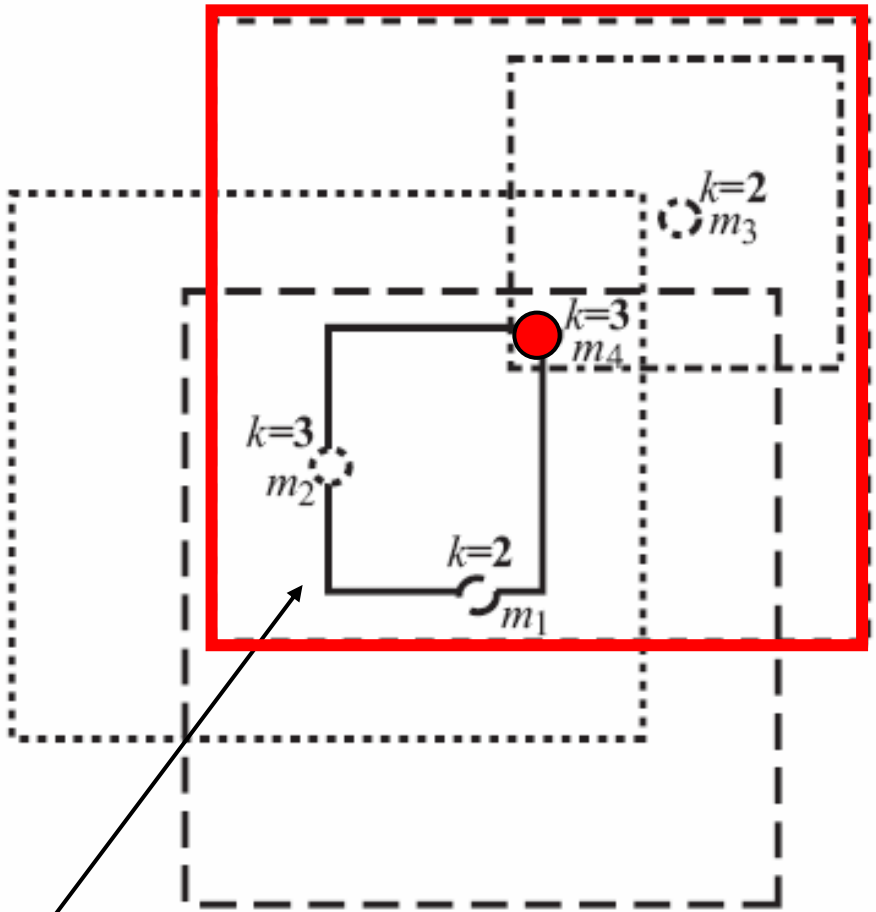


(a) spatial layout I

Clique-Cloak Algorithm: Spatial Layouts



(a) spatial layout I

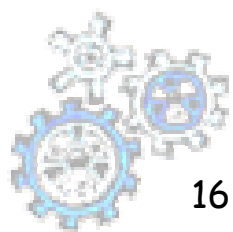


(b) spatial layout II

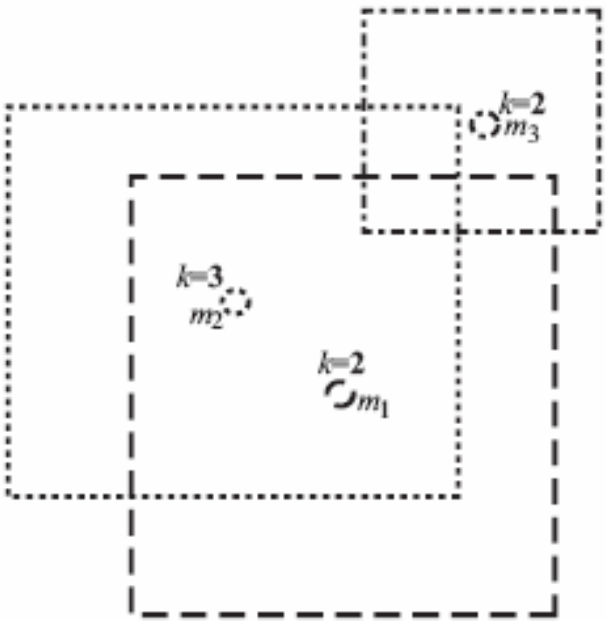
minimum bounding rectangle

Constraint Graphs

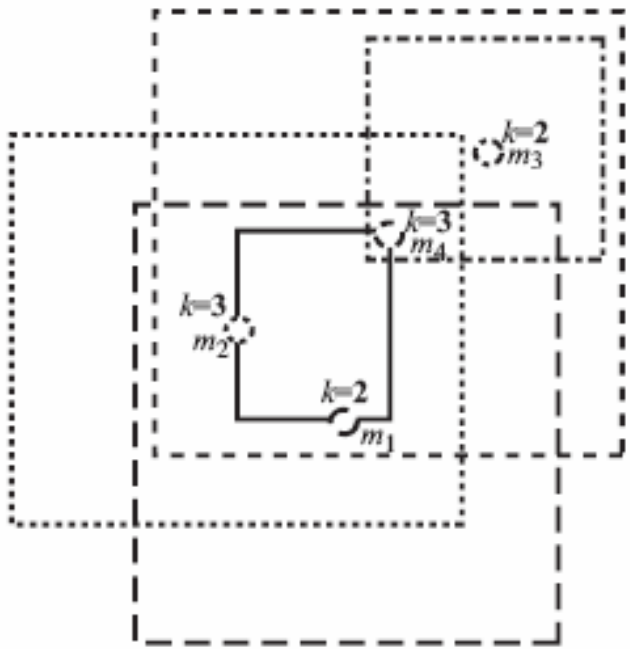
- G(S,E) is an undirected graph
- S is the set of vertices
  - Each representing a message received at the message perturbation engine
- E is the set of edges, $(ms_i, ms_j) \in E$ iff
  1. $L(ms_i) \in Bcn(ms_j)$
  2. $L(ms_j) \in Bcn(ms_i)$
  3. $ms_i.uid \neq ms_j.uid$
- Ms_i is anonymizable iff an $\exists$ an l-clique M s.t. $\forall\ ms_i \in M$ we have $ms_i.k \leq l$



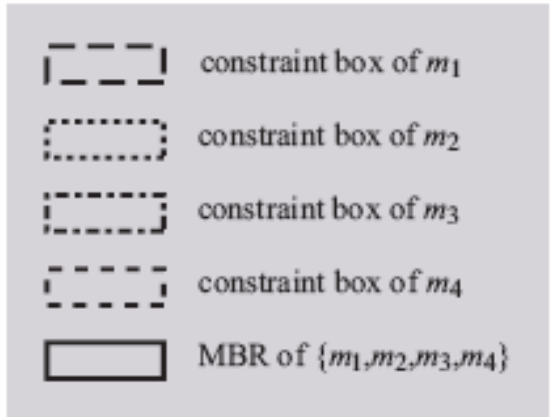
Clique-Cloak Algorithm: Constraint Graphs

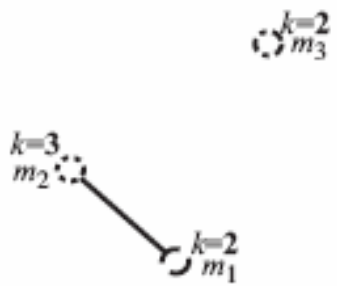


(a) spatial layout I

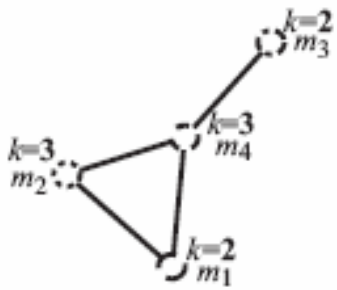


(b) spatial layout II

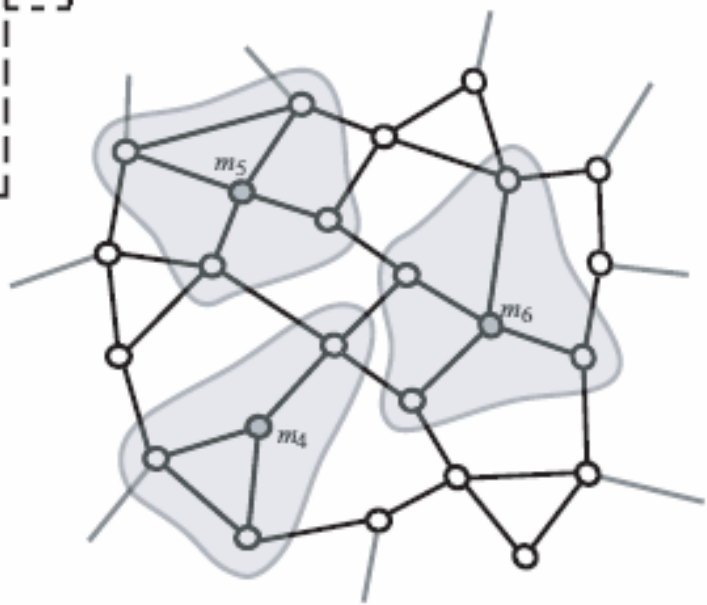




(c) constraint graph I



(d) constraint graph II

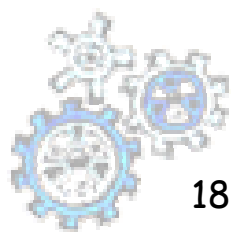


(e) constraint graph

Clique-Cloak Algorithm: Four Steps

- **Data structures: Message Queue, Multidimensional Index, Constraint Graph, Expiration Heap**

- **Steps:**
  1. **Zoom-in,** i.e. Locate neighbors messages of popped message m, update data structures (Index and Graph)
  2. **Detection,** (local k-search sub-algorithm) find a m.k-clique in the subgraph {m} U {$m_j \in$ neighbor of m | $m_j.k \leq m.k$ }
  3. **Perturbation,** use the MBR of the clique as cloaking box of the messages in the clique
  4. **Expiration,** through an expiration heap

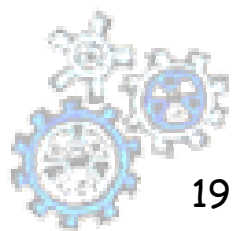


An Optimization: nbr-k Search Algorithm

~~**Detection,** (local k-search) find a m.k-clique in the subgraph of the message and its neighbors m_j s.t. $m_j.k \leq m.k$~~

Detection, (nbr k-search) find the ***largest*** clique M in the subgraph of the message and its neighbors m_j s.t. $m_j.k \leq |M|$

The suggested implementation makes use of local k-search varying k in a decreasing order



Synthetic Data Generator

Parameter	Default value
anonymity level range	$\{5, 4, 3, 2\}$
anonymity level zipf param	0.6
mean spatial tolerance	$100m$
variance in spatial tolerance	$40m^2$
mean temporal tolerance	$30s$
variance in temporal tolerance	$12s^2$
mean inter-wait time	$15s$
variance in inter-wait time	$6s^2$

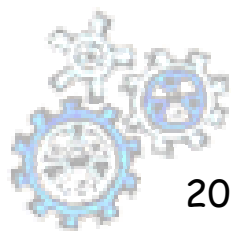
Table 1: Message generation parameters

mean of car speeds for each road type	$\{90, 60, 50\}km/h$
std.dev. in car speeds for each road type	$\{20, 15, 10\}km/h$
traffic volume data	$\{2916.6, 916.6, 250\}$per hour

Table 2: Car movement parameters

Chamblee region of state of Georgia in USA (160km^2)

10,000 cars



Experiments: Success rate and anonymity level

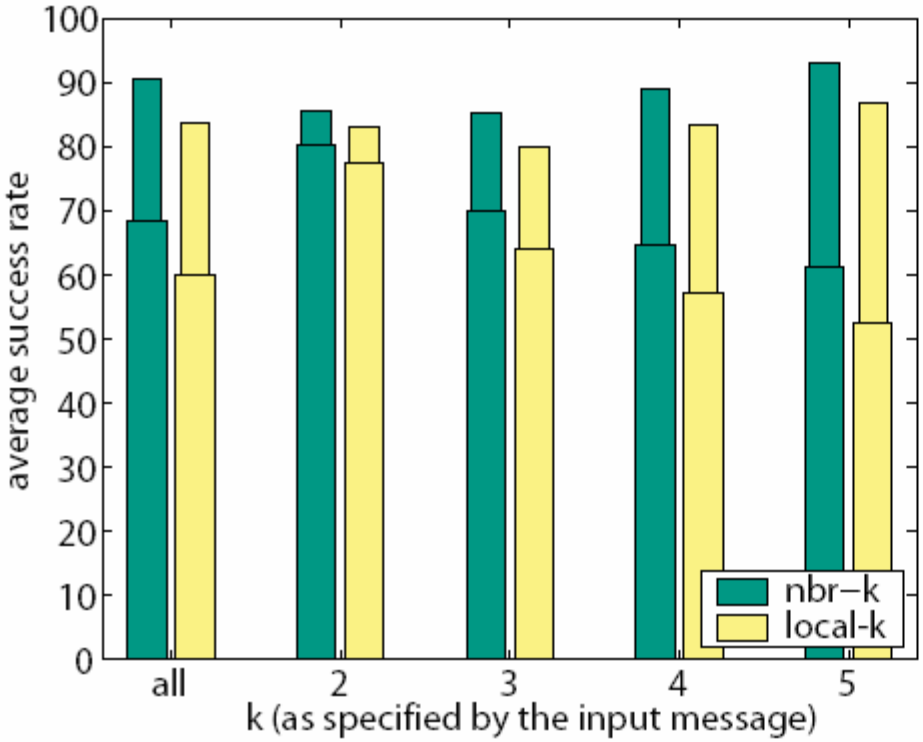


Figure 2: Success rates for different k values

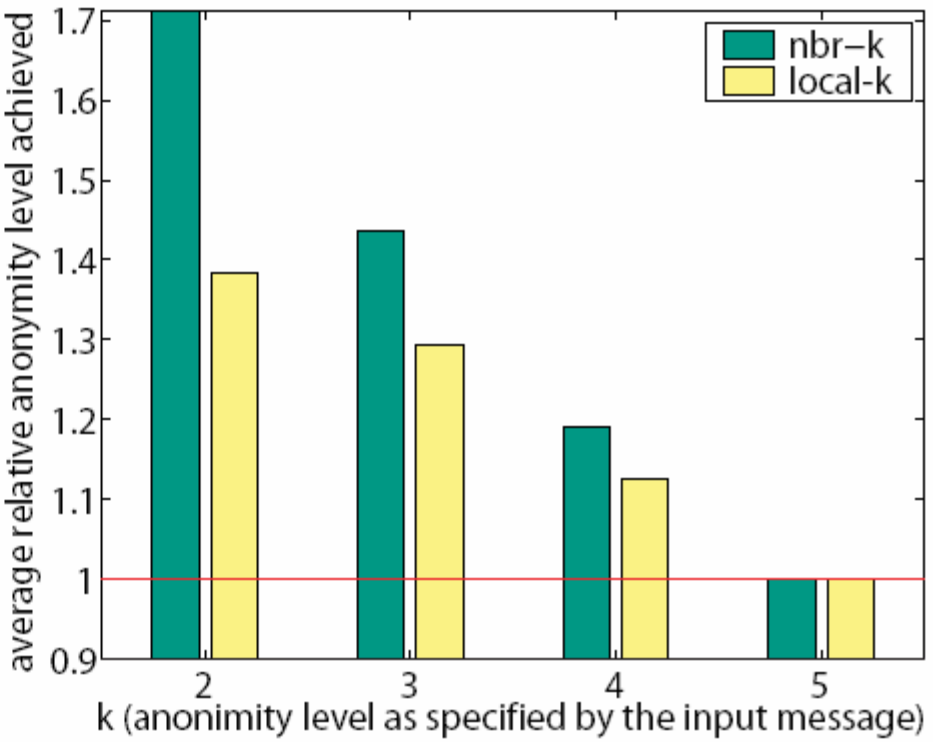


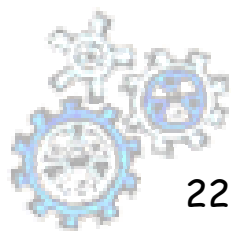
Figure 3: Relative anonymity levels for different k values

Accuracy < 18m in 75% of the cases!

Conclusions

- Need for privacy, expecially for LBS
- K-anonymity seems to be a good model for location privacy

- Only few interesting papers available, work is still in progress!

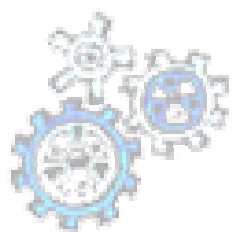


Web Links on Privacy Technologies

- lab.privacy.cs.cmu.edu/people/sweeney/
- www.cs.umbc.edu/~kunliu1/research/privacy_review.html
- www.amstat.org/comm/cmtepc/
- www.cs.ualberta.ca/~oliveira/psdm/psdm_index.html
- www.cs.ut.ee/~helger/crypto/link/data_mining/

- theory.stanford.edu/~rajeev/privacy.html
- theory.stanford.edu/~nmishra/cs369-2004.html

Thank you! ☺



Italian Big Brother Awards 2005

- Negative winners:
  - Telecom – most voted company
  - Dr. Silvio Berlusconi – SMS spamming
  - Avv. Giuseppe Fortunato – Member of Privacy Authority (8 March 2002 guilty of privacy infringements)
  - LazioMatica – Computers used for privacy breaches
  - Miur-Invalsi – large scale survey on students with sensible data
  - Microsoft – Office97 users tracked with UID
- Positive winner:
  - Prof. Stefano Rodotà, ex Privacy Authority